IN THE CLAIMS

Serial No.: 10/600,475

Please cancel claims 30-60 without prejudice.

- (Original) A method for processing audio data, comprising:
 applying a plurality of anchor models to the audio data;
 mapping the output of the plurality of anchor models into frame tags; and
 producing the frame tags;
 wherein the plurality of anchor models comprise a discriminatively-trained
 classifier.
- 2. (Original) The method as set forth in claim 1, wherein the discriminatively-trained classifier comprises a convolutional neural network classifier.
- 3. (Original) The method as set forth in claim 2, further comprising training the convolutional classifier on data separate from audio data available in a use phase.
- 4. (Original) The method as set forth in claim 2, wherein the convolutional neural network classifier is a time-delay neural network (TDNN) classifier.
- 5. (Original) The method as set forth in claim 4, further comprising training the TDNN classifier using cross entropy.
- 6. (Original) The method as set forth in claim 1, further comprising preprocessing the audio data to generate input feature vectors for the discriminatively-trained classifier.
- 7. (Original) The method as set forth in claim 1, further comprising normalizing a feature vector output of the discriminatively-trained classifier.

- 8. (Original) The method as set forth in claim 7, wherein the normalized feature vectors are vectors of unit length.
- 9. (Original) The method as set forth in claim 1, further comprising:
 accepting a plurality of input feature vectors corresponding to audio features
 contained in the audio data; and

applying the discriminatively-trained classifier to the plurality of input feature vectors to produce a plurality of anchor model outputs.

10. (Original) The method as set forth in claim 1, wherein the mapping comprises:

clustering anchor model outputs from the discriminatively-trained classifier into separate clusters using a clustering technique; and

associating a frame tag to each separate cluster.

- 11. (Original) The method as set forth in claim 10, further comprising applying temporal sequential smoothing to the frame tag using temporal information associated with the anchor model outputs.
- 12. (Original) The method as set forth in claim 1, further comprising:
 training the discriminatively-trained classifier using a speaker training set
 containing a plurality of known speakers; and

pre-processing the speaker training set and the audio data in the same manner to provide a consistent input to the discriminatively-trained classifier.

- 13. (Original) A computer-readable medium having computer-executable instructions for performing the method recited in claim 1.
- 14. (Original) A computer-implemented process for processing audio data, comprising:

applying a plurality of anchor models to the audio data;

mapping the output of the anchor models into frame tags; and producing the frame tags;

wherein the plurality of anchor models comprise a discriminatively-trained classifier that is previously trained using a training technique.

- 15. (Original) The computer-implemented process of claim 14, wherein the training technique employs a cross-entropy cost function.
- 16. (Original) The computer-implemented process of claim 14, wherein the training technique employs a mean-square error metric.
- 17. (Original) The computer-implemented process of claim 14, wherein the discriminatively-trained classifier comprises a convolutional neural network classifier.
- 18. (Original) The computer-implemented process of claim 14, further comprising:

omitting an output non-linearity, which was used during training, from the discriminatively-trained classifier to generate a modified feature vector output; and normalizing the modified feature vector output.

- 19. (Original) The computer-implemented process of claim 18, wherein normalizing further comprises creating a modified feature vector output having unit length.
- 20. (Original) A method for processing audio data containing a plurality of speakers, comprising:

applying a plurality of anchor models to the audio data;
mapping an output of the anchor models into frame tags; and
constructing a list of start and stop times for each of the plurality of speakers
based on the frame tags:

wherein the plurality of anchor models comprise a discriminatively-trained classifier previously trained using a training set containing a set of training speakers, and wherein the plurality of speakers is not in the set of training speakers.

- 21. (Original) The method as set forth in claim 20, wherein the discriminatively trained classifier is a time-delay neural network (TDNN) classifier.
- 22. (Original) The method as set forth in claim 20, further comprising normalizing a feature vector output from the convolutional neural network classifier by mapping each element of the feature vector output to a unit sphere such that the feature vector output has unit length.
- 23. (Original) One or more computer-readable media having computer-readable instructions thereon which, when executed by one or more processors, cause the one or more processors to implement the method of claim 20.
- 24. (Original) A computer-readable medium having computer-executable instructions for processing audio data, comprising:

training a discriminatively-trained classifier in a discriminative manner during a training phase to generate parameters that can be used at a later time by the discriminatively-trained classifier:

applying the discriminatively-trained classifier that uses the parameters to the audio data to generate anchor model outputs; and

clustering the anchor model outputs into frame tags of speakers that are contained in the audio data.

25. (Original) The computer-readable medium of claim 24, further comprising pre-processing a speaker training set during the training and validation phase to produce a first set of input feature vectors for the discriminatively-trained classifier.

Serial No.: 10/600,475 Attorney Docket No: MCS-018-03

(Original) The computer-readable medium of claim 25, further comprising 26.

pre-processing the audio data during the use phase to produce a second set of input

feature vectors for the discriminatively-trained classifier, the pre-processing of the audio

data being preformed in the same manner as the pre-processing of the speaker training

set.

27. (Original) The computer-readable medium of claim 24, further comprising

normalizing the feature vector outputs to produce feature vectors having a unit length.

28. (Original) The computer-readable medium of claim 27, wherein normalizing

further comprises omitting a nonlinearity from the discriminatively-trained classifier during

the use phase.

29. (Original) The computer-readable medium of claim 25, further comprising

applying temporal sequential smoothing to the clustering the clustered feature vector

outputs to produce the frame tags.

Claims 30-60: Canceled